鲁棒模式识别研究进展

张煦尧 刘成林
（中国科学院自动化研究所，模式识别国家重点实验室）

摘要

目前诸多模式识别任务的识别精度获得不断提升，在一些任务上甚至超越了人的水平。单从识别精度的角度来看，模式识别似乎已经是一个被解决了的问题。然而，高精度的模式识别系统在实际应用中依旧会出现不稳定和不可靠的现象。因此，开放环境下的鲁棒性成为制约模式识别技术发展的新瓶颈。实际上，在大部分模式识别模型和算法背后蕴含着三个基础假设：封闭世界假设、独立同分布假设、以及大数据假设。这三个假设直接或间接影响了模式识别系统的鲁棒性，并且是造成机器智能和人类智能之间差异的主要原因。本文简要论述如何通过打破三个基础假设来提升模式识别系统的鲁棒性，更详尽的讨论与分析参见[1]。

1. 引言

模式识别方法的演化大致可概括为：统计学习方法、句法结构方法、神经网络与深度学习方法。关于早期模式识别领域的发展历史可参见 1968 年 Nagy[2]、1980 年 Fu[3]、以及 2000 年 Jain 等人[4]的综述论文。从 2006 年[5]开始，深度学习[6]逐渐成为模式识别领域的主流方法。

传统的模式识别方法大都基于人工设计特征结合分类器学习的思想。如图 1 所示，特征提取部分往往是与特定任务相关的（如人脸识别、文字识别、指纹识别等）；而分类器学习（或模式分类）部分则属于更加通用的机器学习问题。与此不同，深度学习将特征提取和分类器学习进行结合，通过端到端的方式，自动地从数据中学习具备更强判别性的特征表示从而实现高精度分类。当前，深度学习已经在不同的模式识别任务上取得优异性能，识别精度被不断刷新。

然而，实验室环境下高精度的模式识别系统一旦部署到真实应用场景中，依旧会出现各种水土不服的现象，即鲁棒性不够。造成模式识别系统不够鲁棒的原因是由

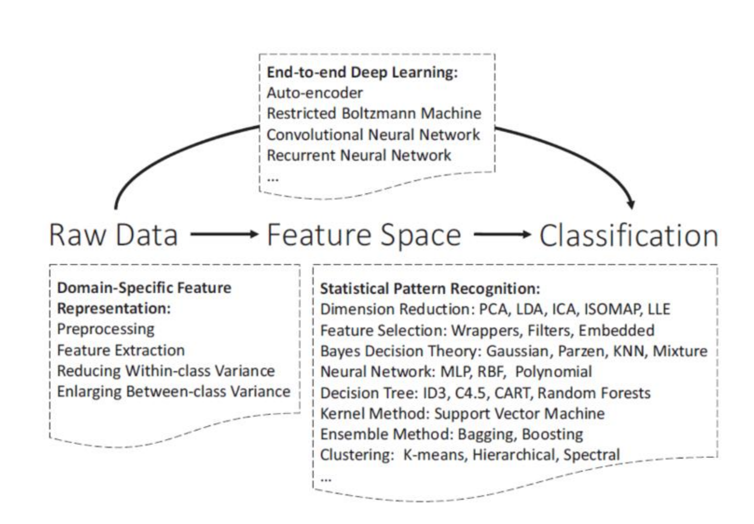


图 1 模式识别基本方法概览

于其背后所蕴含的三个基础假设。其中第一个假设是“封闭世界假设”，即模式识别所处理的类别是事先定义且固定不变的，训练和测试均围绕封闭的 k 个类别展开。在此假设下，分类问题变得更为清晰和明确，只需将特征空间划分成 k 个不同的区域即可。然而，在实际应用中，样本往往来自开放环境，有可能是不属于任何类别的噪声数据，也有可能是训练集未出现过的新类别数据，还有可能是来自混淆区域的对抗数据。在这些情况下，基于封闭世界假设的模式识别系统往往会出现过于自信的明显错误。

第二个假设是“独立同分布假设”，即样本与样本之间是相互独立的，并且训练集和测试集是同分布的。在独立的假设下，模式识别所优化的总损失函数（又叫经验风险）可以转化成每个样本的损失函数之和。而在同分布假设下，则可以预期最大化训练集的精度也能带来最优的测试（泛化）性能。然而，在现实环境中，独立同分布假设往往是不成立的。在各种条件和环境下搜集的数据，不能简单的看成是独立的，非独立数据的上下文关系能有效提升模式识别的鲁棒性。此外，训练集和测试集细微的分布差异就会带来识别性能的大幅下降。

第三个假设是“大数据假设”，即训练数据的规模要

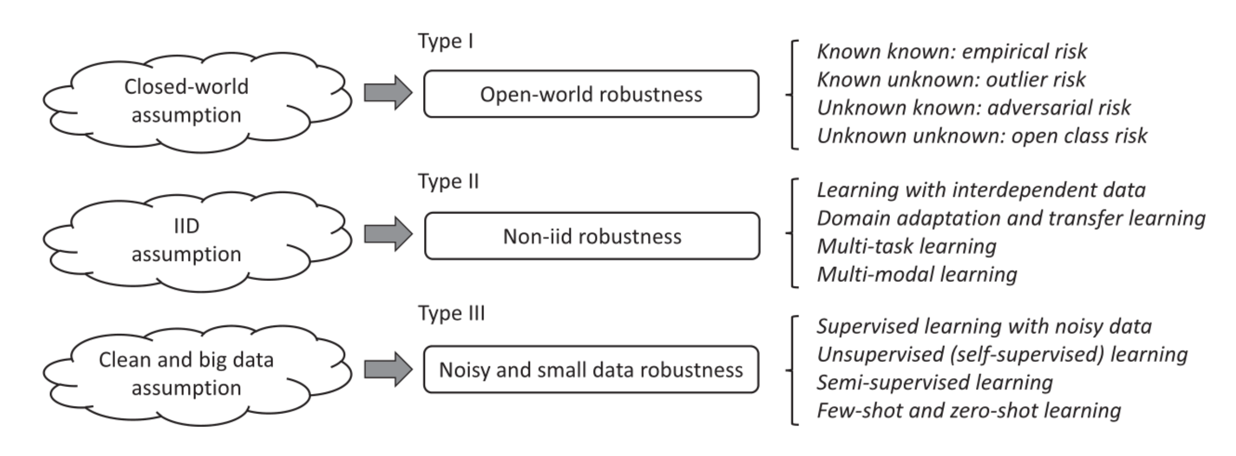


图 2 模式识别中的三个基础假设与鲁棒性问题

足够大以涵盖不同的分布变化，并且数据的标注要尽可能精确。在此假设下，唯一的需求是模型具备足够强大的拟合能力，通过监督学习的方式将获得较好的泛化性能。然而这个假设在真实环境中往往也不成立。首先，对一些特定的模式识别任务，很难去收集大量的训练数据；其次，对所有数据进行精准的标注也是一项很难完成的任务。实际上，如何从少量样本以及弱标记数据中进行有效学习是机器与人类智能之间的显著差异。

针对这种局面，科技部在“科技创新 2030—新一代人工智能重大项目”指南中强调了“面向开放环境的自适应感知”这一研究方向的重要性。同时，在 2016 年国际人工智能发展协会（AAAI）大会上，AAAI 主席 Thomas G. Dietterich 发表了题为“Steps Toward Robust Artificial Intelligence”的主席演讲，也强调了智能系统在开放环境下的自适应性、对噪声和错误的鲁棒性等问题的重要性。实际上，对于开放环境鲁棒模式识别问题的研究，目前各个国家和科研机构都处于起步阶段，所以大家的水平属于“并跑”，在这一领域开展研究并取得突破将有力提升我国在模式识别领域的国际影响力。

如图 2 所示，本文从打破三个基础假设的角度出发，对模式识别中的鲁棒性问题进行分析和讨论。从下文开始，在每个章节（假设）下，分别探讨四类问题，并在最后一节进行总结和展望，希望通过分析当前方法的不足和局限，从而提升模式识别在开放环境中的鲁棒性。

2. 封闭世界假设

大部分模式识别方法均是基于封闭世界假设：尽管只能观测到有限的样本和有限的类别，但大部分模型往往却对特征空间进行完全的划分，如支撑向量机将特征空间通过大间隔的方式划分成两部分、深度神经网络使用 softmax 操作将特征空间划分成固定个数的类别，并且默认类别后验概率之和等于 1。这些封闭世界模型在面对开放环境时会出现鲁棒性欠缺的问题。实际上，在开放的特征空间中存在大量的未知区域，如图 3 所示，为提升鲁棒性，必须有效处理 Known 和 Unknown 的问题。

2.1 Known Known：经验风险

如图 3a 所示，known known 代表“things we know that we know”，即类别是已知的并且每个类别有一定量的已知样本。这也是模式识别问题的传统定义方式。在此设定下，经验风险最小化成为主流的学习方法，即通过最小化训练集上的分类损失来学习分类器。然而，由于训练数据有限，往往会导致过拟合从而使得泛化性能下降。如传统的最近邻分类器往往会过拟合，因此 k 近邻通过在验证集上寻找合适的 k 来改善泛化性；决策树方法[7]需要通过剪枝等手段来防止过拟合；多层神经网络模型[8]可以拟合任意复杂的函数，因此需要使用不同的策略来防止过拟合。

结构风险最小化[9]通过在经验风险和模型复杂度之间寻找折中来改善泛化性能，如支撑向量机[10]中使用的大间隔正则项可以有效的防止过拟合。此后，很多其他正则项也被广泛使用如稀疏正则[11]、低秩正则[12]、流形正则[13]等。一些隐式的操作也可看成是特殊的正则项如带噪声训练[14]、dropout[15]等。针对传统的 Known Known 问题，经验风险最小化结合正则项的方法被广泛采用。

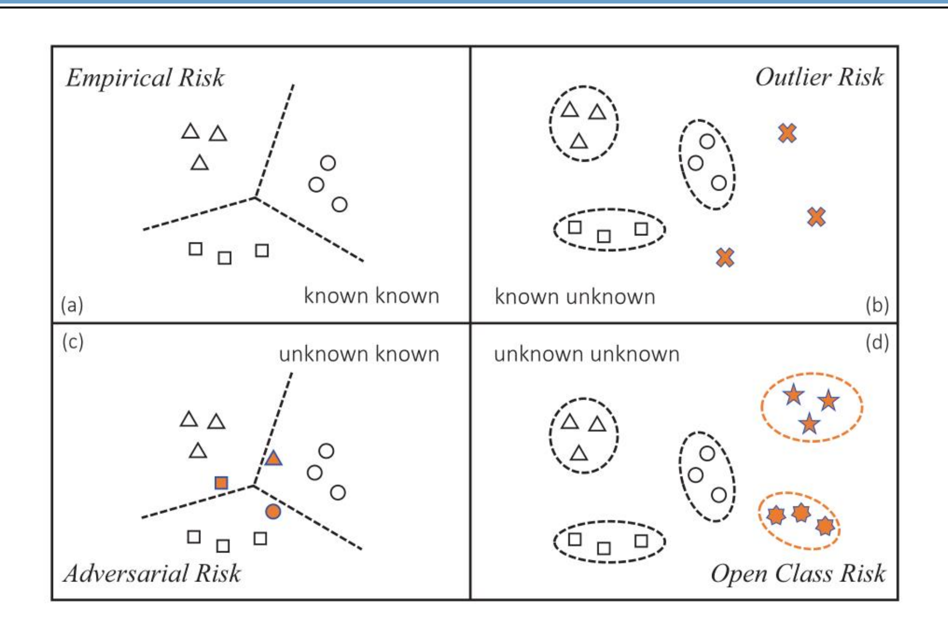


图 3 开放空间的四种情况：(a) known known, (b) known unknown, (c) unknown known, (d) unknown unknown

2.2 Known Unknown：噪声风险

如图 3b 所示，known unknown 代表“we know there are some things we do not know”。在开放环境中，样本有可能是不属于任何类别的噪声数据。处理噪声数据最直接的方式是将传统的 k 类问题扩展到 k+1 类，即用额外的一个类别来表示噪声。然而，这一做法需要收集噪声数据用于训练，由于噪声分布的多样性，很难无偏地采集数据，所以将导致噪声类别难以建模和泛化。

常用的解决方案是模式拒识，即通过给传统的分类器设计相应的拒识规则来屏蔽噪声数据。如贝叶斯分类器的拒识[16]、支撑向量机的拒识[17]、最近邻分类器的拒识[18]、稀疏表示的拒识[19]等。文献[20]指出：不同的分类器结构和学习算法对拒识性能有较大影响。因此，需要针对不同的分类器类型设计相应的拒识准则。

Softmax 是神经网络中常用的分类层函数，可以看成是不同类别的后验概率。一般的做法是针对最大的概率值或者最大的两个概率值之差，设定阈值来进行拒识。然而由于 softmax 的封闭世界性质（概率和为 1），很难取得满意的拒识效果。一个有效的改进是采用 sigmoid 函数和 one-vs-all 训练方式[21][22]来改善拒识性能。对 softmax 函数的其他改进还包括：openmax [23]、generative openmax [24]等。

基于 one-class 思想的方法将所有数据看成一类，从而实现对未知数据的拒识，代表性工作有[25]和[26]等。文献[27]从理论上定义和讨论了如何在分类器的训练过程中考虑噪声风险，而文献[28]利用统计理论方法来处理开放集识别问题。“学会拒绝”是封闭世界和开放世界模式识别的首要区别，虽然有不少工作关注这一问题，然而更加简洁高效的方法仍然需要不断探索。

2.3 Unknown Known：对抗风险

如图 3c 所示，unknown known 代表“things we think we know but it turns out we do not”。这一现象往往发生在不同类别分界面附近的易混淆区域，由于有限的训练样本很难覆盖这一区域，导致这个区域的样本容易被错误分类。实际上，真实世界中样本落在这个区域的频率也是很低的。然而，研究人员从算法的角度人工生成此类数据，称之为“对抗样本”[29]，对模式识别的鲁棒性提出了严峻的挑战。

通过给图像增加一个细微的肉眼几乎不可见的扰动，可以彻底改变模式识别系统对该图像的类别预测，说明较小的输入端扰动带来了较大的输出端变化。对抗样本可以通过利用梯度信息[30]或者优化的方式[31]来获得，只有图像像素千分之一大小的扰动就足以欺骗主流的深度学习系统。利用对抗样本来攻击模式识别系统将在一些对安全性要求较高的应用中造成较大的风险。

因此，有必要针对对抗样本设计相应的防御策略。一个主流的做法是将对抗样本作为增广的数据[30]用于训练来提升系统的鲁棒性，或利用一个检测器来自动区分对抗样本和正常样本[32]。梯度平滑[33]和鲁棒优化

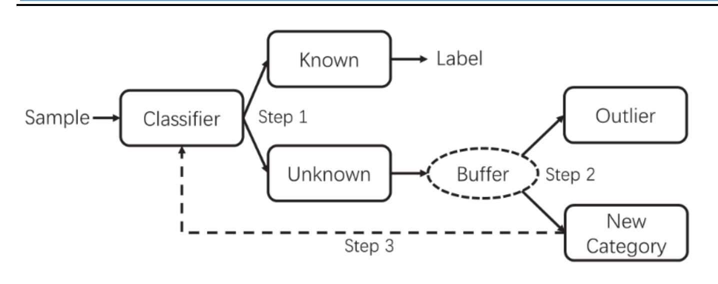


图 4 开放环境类别增量学习

[34]也被用来防御对抗样本。如图 3c 所示，对抗样本往往出现在易混淆的区域，因此可以通过增加该区域的训练数据来提升鲁棒性，如文献[35]和[36]通过不同类别样本的线性插值来人为地构造新数据，一定程度上模拟了分界面附近的数据分布。目前，各式各样的对抗攻击方法依旧在不断地被提出，因此有效的防御手段和鲁棒学习方法是模式识别面临的重要研究课题。

2.4 Unknown Unknown：开放类别风险

如图 3d 所示，unknown unknown 代表“unknown samples grouped into unknown classes”。在模式识别系统的实际应用中，数据往往是连续不断出现的，并且类别数也是动态变化的，这一现象在学术界被称作开放世界识别[37]或类别增量学习[38]。

如图 4 所示，类别增量学习的三个基本步骤是：第一步是模式拒识，即判断该样本是已知的还是未知的，如果是已知样本则进行识别，如果是未知样本则直接拒识，将其暂存于一个寄存器之中；待寄存器中积累了足够数量的样本之后，第二步是新类别的发现，要判断这些被拒识的样本到底是噪声还是属于特定的新类别；在得到新类别以及相应样本之后，第三步则是类别增量学习，对模式识别系统中存储的类别进行动态扩充。

第一步模式拒识在 2.2 节已有介绍。第二步新类别发现，可以通过对寄存器中的样本进行聚类分析[39]来实现：如果某一个聚类中包含足够的样本量则可以视为一个新的类别，如果某一个聚类只有零星的几个样本则可以视为噪声而忽略掉。用聚类的方式来自动发现新类别的难点在于如何确定新类别的个数，因此聚类模型要具备自动的模型选择能力[40]来确定聚类中心的个数。

有了新类别及其相应的样本之后，最后一步则是要对整个模式识别系统进行调整使得其具备新类别的识别能力。一般来说，在模式识别问题中，判别式模型具有更高的识别精度，而生成式模型更适合类别增量任务。因此，可以结合二者的优势设计算法，如文献[41]在学习一个判别子空间的同时，利用最近中心分类器几乎实现了零成本的新类别扩增。同样的思想也可以用于深度学习，如文献[42]提出了深度最近中心分类器、文献[43]提出了卷积原型网络等。

然而，类别增量学习的另外一个问题是：当新类别加入时如果只调整新类别判别函数，在新类上的精度会较低，而如果同时调整特征表示以及判别函数，由于旧类别的数据此时是缺失的，会带来“灾难性遗忘”问题，即新类别的识别率提升了而旧类别的识别率却大幅下降了。为了解决这一问题，文献[38]提出增量表示和分类器协同学习的框架 iCaRL 通过保存一部分旧类别代表性数据来克服遗忘，而文献[44]利用新类别数据在旧类别上进行知识蒸馏来一定程度上弥补遗忘问题。如何克服灾难性遗忘是连续学习或类别增量学习面临的严峻挑战，近年来吸引了大量研究者关注。

3. 独立同分布假设

独立同分布是模式识别中的重要基础假设。在 2015 年举办的国际研讨会[45]上，参会学者一致认为如何有效地从非独立和非同分布数据中进行学习是一项重要而富有挑战的课题。文献[46]揭示了微小的分布变化会导致模式识别系统显著的性能下降。如图 5 所示，根据输入特征空间和输出类别空间是否改变，可以将非独立同分布的情形划分成不同的任务，下面分别展开论述。

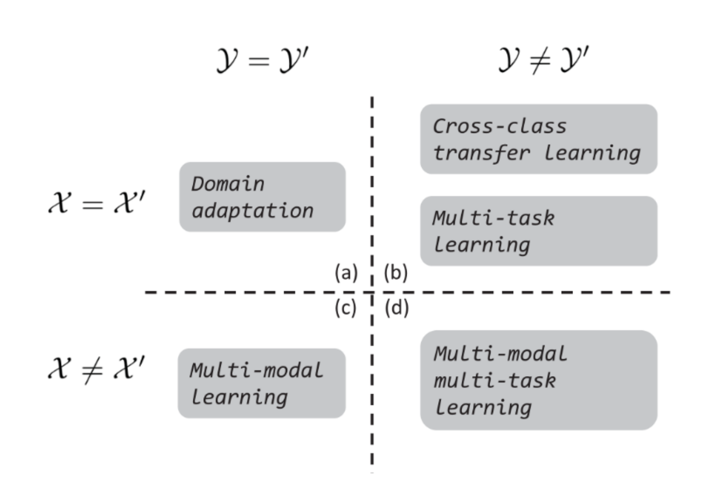


图 5 不满足独立同分布假设的各种任务

3.1 非独立数据学习

传统模式识别往往假设数据之间是独立的。但是在真实环境中，数据往往会“成组”出现（在文献中又被称为 group，set，bag，或者 field）。这种“组”信息提供了样本之间的上下文依赖关系：在每一个组内部样本不再是独立的，而不同的组之间也不一定是同分布的。充分利用这种约束信息将极大提高决策的鲁棒性。

第一种情形是“内容一致性”即一个组内的所有样本均来自同一个类别。这一问题在学术界又被称为 image set [47]或 group-based [48]分类。相比于单一样本的决策，同一个组里的不同样本将从不同的角度(如姿态、光照、形变等)提供互补信息从而提升决策准确性。第二种情形是“风格一致性”即一个组内的所有样本均具备一致的特定风格，如同一个人书写的若干文字、同一视角下拍摄的多幅图像、同一种口音说出的语音信号等。这在学术界又被称为 pattern field classification [49][50]。同源样本之间的风格一致性是提升决策鲁棒性的有效途径。第三种情形是多示例学习[51]，即组内的单个样本是没有标记的，只在组层面提供标记信息。在此设定下，组内样本不再是独立同分布的[52]，并为实现弱监督学习提供了有效途径。

成组样本的排列关系即上下文信息[53]也是提升模式识别鲁棒性的重要手段，如语言上下文、几何上下文等。一种有效利用上下文关系的方式是将不同样本看成序列，然后利用隐马尔可夫模型[54]、条件随机场[55]、递归神经网络[56]等来对序列数据进行建模。图也是关系表示的重要方式，因此，图神经网络[57][58]在上下文建模中也越来越重要。此外，结构化预测[59]可以充分利用输出标签之间的依赖关系来提升预测准确性。

3.2 自适应与迁移学习

当训练数据（源域）和测试数据（目标域）分布发生变化时，模式识别性能会大幅下降，因此自适应和迁移学习变得尤为重要。当目标域中拥有一定的标记样本时，最简单直接的方式是对源域训练好的模型进行微调，称之为有监督的自适应。另外一类方法“跨域映射”，通过学习源域[60]或者目标域[61]的映射函数，来消除两个域之间的分布差异，既可以用于有监督也可以用于无监督自适应。此外，还可以通过对样本进行重加权[62]的方式来消除分布差异。不同的度量方式（如分布距离、散度、信息量等）[63]可以用来有效衡量分布差异从而实现自适应。

近期，对抗学习[64][65]也被广泛用于自适应和迁移学习。其基本思想是尽可能地让两个域的数据分布差异无法区分，从而得到域不变的特征表示。通过利用“基分类器”和“域分类器”的相互对抗来实现，基分类器实现的是传统分类任务，而域分类器则是要判断数据是来自源域还是目标域，二者共享底层的特征提取。对抗的结果是源域和目标域的数据无法区分，从而消除两个域之间的分布差异。

在模式识别问题中通过自适应可以显著改善识别性能，如文字识别中的书写人自适应，语音识别中的说话人自适应，图像分类中的视角光照自适应等。传统的自适应往往假设只有一个源域，然而在实际应用中，多源问题[66]经常发生，需要将算法进行改进以满足多个源域的需求。此外，当多源数据混杂在一起时，如何自动地发现其中隐含的域[67]也是影响自适应性能的关键因素。

3.3 多任务学习

在模式识别中，同一输入信号实际上可以用来完成多种任务，如一张人脸图像可以用来预测种族、年龄、性别等。这些任务相互之间不是独立的，且由于输出类别的不一致导致其分布也是不同的。充分利用多种任务之间的互补信息可以达到分别提升彼此性能的效果。

多任务学习考虑的第一个问题是能否找到一个通用的特征表示，在不同的任务之间较好地迁移。传统的人工特征显然是与任务相关的，而深度神经网络预训练好的特征可以适用于不同的任务[68]。由于深度神经网络的分层结构，其底层往往学习的是一些低级特征而高层则是更强的语义特征。底层特征更加一般化而高层特征更与任务相关，因此高层特征更适合在相似任务间迁移，而底层特征更适合在差异较大的任务间迁移[69]。

多任务学习考虑的第二个问题是能否结合多个任务的监督信息来学习不变的特征表示。实际上，当一个模型具有多个损失函数时，很大程度上即是在进行多任务学习。在此过程中，如何设计好“任务共享”与“任务相关”的参数是问题的核心。一个简单直接的方式是不同任务共享底层特征，而高层决策则是与任务相关的[70]。为更好的实现任务之间的交互，也可以通过约束的方式[71]或学习的方式[72]来实现更加自由的信息共享。

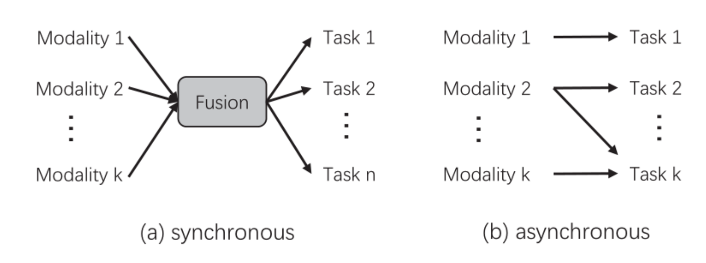


图 6 两种不同的多模态多任务学习

多任务学习中的另外一个重要问题是如何学习任务之间的相互关系[73]。定义好任务之间的关系可以大大提升迁移学习的性能。对于特定的任务，可以只在与之最相关的任务之间进行迁移。和使用所有任务相比，在降低计算量的同时也能避免不相似任务在迁移过程中造成的负面影响。此外，如何均衡不同任务损失对训练过程的影响也是多任务学习中的重要问题，可以通过学习不同任务权重[74]以及对不同任务梯度进行归一化[75]的方式来更好地学习多任务模型。

3.4 多模态学习

利用多模态信息来提升模式识别鲁棒性的例子非常普遍，如生物特征识别中可以融合人脸、指纹、虹膜等来实现更精准的预测，无人驾驶中可以融合雷达、摄像机、GPS 等信号来实现更加鲁棒和安全的决策等。

多模态学习的第一个问题是如何设计有效的多模态融合策略。由于不同模态数据的异构特性，很难在原始数据层面进行融合。主流的方法是在“特征层面”[76]对多模态数据进行融合，或者在“决策层面”[77]对多模态信息的预测结果进行融合。由于深度学习的分层表示机制，在“中间层”进行多模态融合也逐渐受到关注[78]。通过多模态融合，能有效提升模式识别的准确性。然而，在实际应用中，如何有效考虑“模态缺失”[79]（即有些样本可能不包含某些模态）是一个关键问题。

跨模态学习[80]也逐渐受到学术界的关注。第一个例子是“跨模态检索”[81]，通过将不同模态数据映射到相同语义空间，可以利用一种模态去检索另一种模态。第二个例子是“跨模态配准”，对不同模态数据的子模块之间进行配准，如将一部电影中的镜头（图像）与剧本章节（文本）进行对齐[82]等。深度学习中的注意力机制是实现跨模态配准的有效手段。第三个例子是“跨模态生成”，即利用一种模态的数据去自动生成另一种模态的数据，如从图像生成其文本描述[83]等。跨模态学习需要对不同模态内部的交互机制进行探索和建模。

对于模式识别而言，多模态实际上增加了输入端的多样性，而多任务则是增加了输出端的多样性。在实际系统中，多样性往往会带来鲁棒性。因此，多模态多任务学习对于鲁棒模式识别尤为重要。如图 6 所示，多模态多任务学习分为两种情况，第一种情况是“同步”即所有模态对于每一个任务均是有效的，此时可以通过多模态融合后再结合多任务训练来提升系统的整体鲁棒性。然而，一个较为困难的设定是图 6b 所示的“异步”多模态多任务学习，如图像分类处理的对象是图片，语音识别处理的对象是声音，而机器翻译处理的对象是文本。直观上，因为它们的输入和输出均不相同，很难将这些问题联合考虑，并且将之联合考虑到底有没有益处也很难确定。文献[84]为我们展示和验证了这种可能性，体现出多模态多任务学习的巨大潜力。

4. 大数据假设

以深度学习为代表的模式识别系统具备强大的训练数据拟合能力。如文献[85]所示，即便将训练数据的标签随机打乱，神经网络依旧能够取得很小的训练误差。当我们拥有一个规模较大且标注精准的数据库时，较好的训练数据拟合将同时带来较好的泛化性能。然而，在实际问题中，“大数据”和“精准数据”往往是矛盾的：对较小的数据库进行逐个样本的精准标注是可以实现的，然而对于大数据的搜集将不可避免地存在噪声和错误数据。因此，为提升模式识别在数据的“量”和“质”方面的鲁棒性，必须从如下的四个方面展开研究。

4.1 数据容错学习

大数据的搜集将不可避免地带来错误数据，表现为三种形式。第一是标签错误即样本本身是正确的但由于人工标注而引入标签误差，第二是样本错误即由于样本本身被污染而带来的偏差如图像内容被遮挡或发生形变等，第三是噪声错误即样本是不属于任何预定义类别的无意义数据，但仍旧被标注成了其中某一类。

针对数据容错学习第一大类方法是改良损失函数。在传统模式识别中，为了追求较好的最优解，损失函数往往被设计成凸函数。但是，凸函数往往是无界的，错误数据会占据较大的损失从而支配训练过程。因此，鲁棒

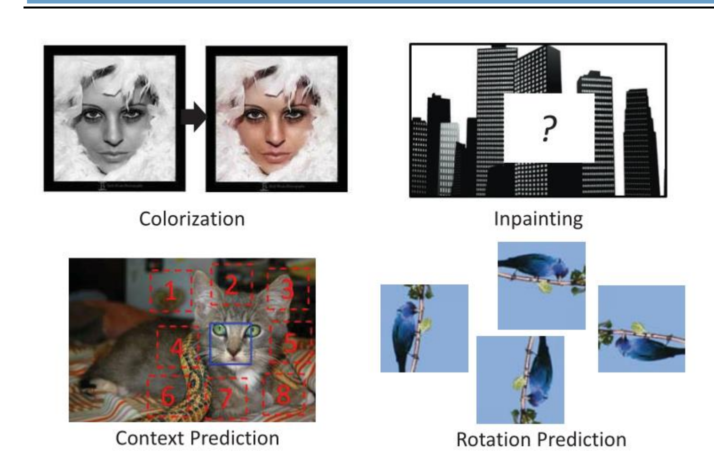


图 7 不同的图像自监督任务

的损失函数往往是非凸的，例如改良后的 ramp loss [86]和 truncated hinge loss [87]等。文献[88]比较了不同损失函数在噪声和错误数据上的处理能力。

在数据的标注过程中，尤其是针对大类别集问题（如汉字识别等），有些类别之间本身就很容易混淆，因此极易造成标记错误。可以通过估计一个标记转移矩阵[89]来衡量两个类别之间被错误标记的概率，从而在所有可能的潜在类别上计算损失来提升训练的鲁棒性。此外，也可以在训练的过程中自动检测和删除错误数据，如通过集成学习的方法[90]或将错误数据的检测嵌入到目标函数的优化过程中[91]等。对数据进行重加权从而赋予错误数据较小的权重[92]也是一种重要的训练方法。此外，还可以通过在训练的过程中对错误数据的标记进行自动更正[93]来提升系统的鲁棒性。

4.2 无监督（自监督）学习

在传统模式识别中，无监督学习往往指的是数据聚类[4]。随着深度学习的兴起，研究的重心慢慢转移到无监督表示学习[94]，即利用无标记数据学习一个通用和可迁移的特征表示。其中自监督学习（即通过数据本身构造监督目标）成为无监督学习的一种新手段。有如下一些方法。

第一类方法是基于重构的方法，即通过特定的编码-解码网络结构对无标记数据进行重构。早期的主成份分析（PCA）模型实际上即是基于这种思想，后来的受限玻尔兹曼机[5]和自编码机[95]也是基于重构的方法并可以看作是 PCA 模型的非线性扩展。关于自编码机，后续衍生出很多改进算法，通过自己重构自己的方式来学习较好的语义特征表示。

既然数据是无标记的，那么一个直观的想法是能否给数据设定相应的“伪标签”，从而将问题转化成传统的监督学习问题。事实上，可以通过聚类算法将数据划分成不同的组，再将组的 id 作为该样本的伪标签来进行监督学习。然而问题的难点在于聚类算法的成功依赖于好的特征表示，而特征表示的学习又需要聚类结果作为监督。因此，需要利用轮替学习的方式来同时进行聚类和特征表示学习[96]。此外，还可直接将样本的 id 当作伪标签，即通过训练使得所有的样本都尽可能分开。此时类别数等于样本数，如何克服大类别数造成的影响是问题的关键[97][98]。

还可以通过设定一系列“辅助任务”的方式来进行自监督学习。如图 7 所示，在视觉任务中可以采用的辅助任务包括：灰度图像彩色化、图像空缺恢复、图像块关系预测、图像旋转角度预测等。对于这些任务，其监督目标可以通过算法自动生成，因此可以利用海量无标记样本开展训练。此外，为完成这些任务，学习到的特征将对图像内容以及空间关系具有较好的表征，从而具备较好的迁移性能。此外，在自然语言处理领域[99]，自监督学习也取得了优异的性能表现。由于不同的自监督任务是从不同的角度提出的，通过多任务学习（3.3 节）的方式将其进行整合[100]可以获得更为丰富的特征表示。

4.3 半监督学习

通过结合无监督学习和监督学习，半监督学习可以在使用少量标记样本和大量无标记样本的情况下进行有效学习，其基本思想是让模型的预测在样本构成的流形上尽可能平滑来改善分类性能。

将传统的监督损失与无监督损失进行结合即可进行有效的半监督学习，如文献[101]将自编码机的重构损失与分类损失结合用于半监督学习。Self-training [102]是一种简单有效的半监督方法，通过轮替的方式让模型给无标记数据预测伪标签再利用其进行监督学习；Co-training [103]利用两个不同的模型在无标记数据上为彼此提供伪标签用于训练；而 tri-training [104]利用 bootstrap 得到三个不同的模型在无标记数据上相互指导从而实现半监督学习。

在深度学习领域，生成对抗网络通过生成器和判别器的相互对抗实现半监督学习[105]，一方面可以生成更

多样本进行数据扩充，另一方面即便是生成的 bad example [106]依旧能提升半监督学习性能，因为它们往往位于流形的低密度区域可以让分类器更好地调整分界面。深度学习的另一个特点是其运行过程中往往会有一定的随机性（如 dropout 或者数据扰动等），因此即便是同一样本两次输入同一网络，其输出也可能不一致。通过最小化这种不一致性，可以提升网络在无标记数据上的平滑性从而用于半监督学习[107]。

从“老师-学生”的角度也能实现半监督学习，如文献[108]利用训练历史过程中集成的预测结果作为“老师”来定义“学生”网络在无标记数据上的训练目标；而文献[109]将“老师”网络的参数定义为“学生”网络参数的历史平均，然后约束无标记数据在老师和学生网络之间的输出差异尽可能小。这实际上是在网络训练的时序过程中保持了无标记样本的预测一致性，实现了局部平滑，从而提升半监督学习性能。在只有少量标记样本并可充分利用无标记样本的情形下，半监督学习是提升分类性能的有力手段。

4.4 小样本与零样本学习

对于人类智能而言，我们仅仅需要观察很少量的样本就可以快速地学会一个新的概念。但是，模式识别中的主流模型往往都是严重依赖大数据的，因此小样本[110]甚至零样本[111]学习变得尤为重要。

人脑具备很强的小样本学习能力，但是这种能力并不是凭空而来的，而是在连续不断的学习过程中积累学习经验之后慢慢形成的。因此，如图 8 左所示，小样本学习的核心思想是：通过在大数据上（many-shot，每个类别可以拥有较多样本）进行学习，不断积累经验后，将这种学习的能力迁移到小样本新类别上（few-shot，每个类只有很少量样本）。并且 many-shot 和 few-shot 数据集所对应的类别之间是没有交集的，所以这一过程实际上可以理解为“小样本的跨类别迁移学习”。

小样本学习的一个极端情况是零样本学习，即新的类别完全没有训练样本。在这种情况下，为了实现知识的有效迁移，需要利用一些辅助信息如类别属性、类别名称、类别文本描述等。实际上，人脑之所以具备零样本学习能力，是因为我们有其他的资源（如互联网、书籍等）从中我们可以推测和顿悟出新类别会是什么样的。所以，如图 8 右所示，零样本学习的本质更像是跨模态

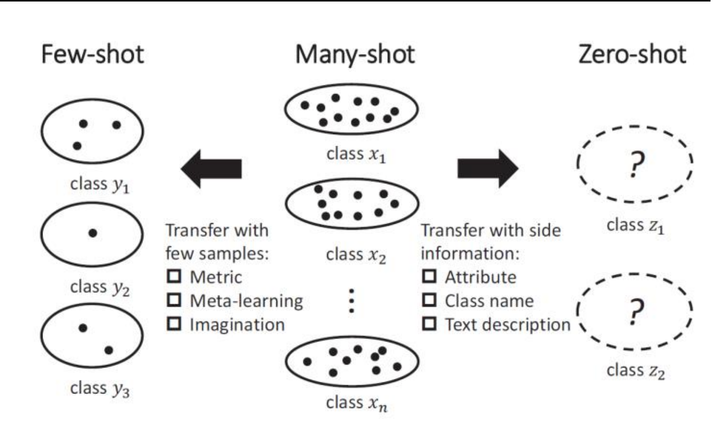


图 8 小样本和零样本学习基本原理图

学习（3.4 节），即从其他模态获取信息来解决当前模态某些类别样本缺失的问题。

目前，在每年的顶级会议和期刊上涌现出很多小样本和零样本学习的前沿工作，然而它们的性能水平相比大数据的模式识别仍然有很大差距，因此是模式识别领域有待研究的挑战性问题之一。

5. 总结与展望

本文简要论述了如何通过打破三个基础假设来提升模式识别的鲁棒性。模式识别的研究不能仅仅局限在提高识别精度上，当更多的评价指标[1]被考虑进来时可以发现模式识别领域还有很多亟待解决的科学问题值得研究。此外，三个基础假设之间也不是孤立的，通过联合思考，很多新的研究问题也会应运而生。通过消除三个假设的影响，鲁棒模式识别的研究将促进相关技术更好地应用于开放环境中的实际问题。

【参考文献】

[1] X.-Y. Zhang, C.-L. Liu, and C. Y. Suen. Towards robust pattern recognition: A review. Proceedings of the IEEE, vol. 108, no. 6, pp. 894-922, June 2020.

[2] G. Nagy. State of the art in pattern recognition. Proceedings of the IEEE, 56(5):836–863, 1968.

[3] K.-S. Fu. Recent developments in pattern recognition. IEEE Trans. Comput., 29(10):845–854, 1980.

[4] A.K. Jain, R.P.W. Duin, and J. Mao. Statistical pattern recognition: A review. IEEE Trans. Pattern Anal. Mach. Intell., 22(1):4–37, 2000.

[5] G. Hinton and R. Salakhutdinov. Reducing the dimensionality of data with neural networks. Science, 313(5786):504–507, 2006.

[6] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. Nature, 521(7553):436–444, 2015.

[7] L. Breiman. Classification and Regression Trees. Routledge, 2017.

[8] C. Bishop. Neural Networks for Pattern Recognition. Oxford university press, 1995.

[9] V.N. Vapnik. Statistical Learning Theory. New York: John Wiley & Sons, 1998.

[10] C. Cortes and V. Vapnik. Support vector machine. Machine Learning, 20(3):273–297, 1995.

[11] T. Zhang. Analysis of multi-stage convex relaxation for sparse regularization. J. Mach. Learn. Res., 11:1081–1107, 2010.

[12] W. Dong, G. Shi, X. Li, Y. Ma, and F. Huang. Compressive sensing via nonlocal low-rank regularization. IEEE Trans. Image Process., 23(8):3618–3632, 2014.

[13] Z. Zhang and K. Zhao. Low-rank matrix approximation with manifold regularization. IEEE Trans. Pattern Anal. Mach. Intell., 35(7):1717–1729, 2013.

[14] C. Bishop. Training with noise is equivalent to Tikhonov regularization. Neural Computation, 7(1):108–116, 1995.

[15] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. J. Mach. Learn. Res., 15(1):1929–1958, 2014.

[16] C. Chow. On optimum recognition error and reject tradeoff. IEEE Trans. Information Theory, 16(1):41–46, 1970.

[17] Y. Grandvalet, A. Rakotomamonjy, J. Keshet, and S. Canu. Support vector machines with a reject option. In Advances Neural Inf. Process. Syst., pages 537–544, 2009.

[18] P. Junior, R. Souza, R. Werneck, B. Stein, D. Pazinato, W. Almeida, O. Penatti, R. Torres, and A. Rocha. Nearest neighbors distance ratio open-set classifier. Machine Learning, 106(3):359–386, 2017.

[19] H. Zhang and V. Patel. Sparse representation-based open set recognition. IEEE Trans. Pattern Anal. Mach. Intell., 39(8):1690–1696, 2017.

[20] C.-L. Liu, H. Sako, and H. Fujisawa. Performance evaluation of pattern classifiers for handwritten character recognition. Int. J. Document Anal. Recognit., 4(3):191–204, 2002.

[21] C.-L. Liu. One-vs-all training of prototype classifiers for pattern classification and retrieval. In Int. Conf. Pattern Recognition, pages 3328–3331, 2010.

[22] L. Shu, H. Xu, and B. Liu. DOC: Deep open classification of text documents. In Conf. Empirical Methods in Natural Language Processing, pages 2911–2916, 2017.

[23] A. Bendale and T. Boult. Towards open set deep networks. In IEEE Conf. Comput. Vis. Pattern Recognit., pages 1563–1572, 2016.

[24] Z. Ge, S. Demyanov, Z. Chen, and R. Garnavi. Generative openmax for multi-class open set classification. arXiv:1707.07418, 2017.

[25] D. Tax and R. Duin. Support vector data description. Machine Learning, 54(1):45–66, 2004.

[26] B. Scholkopf, J. Platt, J. Shawe-Taylor, and A. Smola. Estimating the support of a high-dimensional distribution. Neural Computation, 13(7):1443–1471, 2001.

[27] W. Scheirer, A. Rocha, A. Sapkota, and T. Boult. Toward open set recognition. IEEE Trans. Pattern Anal. Mach. Intell., 35(7):1757–1772, 2013.

[28] W. Scheirer, L. Jain, and T. Boult. Probability models for open set recognition. IEEE Trans. Pattern Anal. Mach. Intell., 36(11):2317–2324, 2014.

[29] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus. Intriguing properties of neural networks. In Int. Conf. Learn. Representations, 2014.

[30] I. Goodfellow, J. Shlens, and C. Szegedy. Explaining and harnessing adversarial examples. In Int. Conf. Learn. Representations, 2015.

[31] S. Moosavi-Dezfooli, A. Fawzi, and P. Frossard. DeepFool: a simple and accurate method to fool deep neural networks. In IEEE Conf. Comput. Vis. Pattern Recognit., pages 2574–2582, 2016.

[32] J. Metzen, T. Genewein, V. Fischer, and B. Bischoff. On detecting adversarial perturbations. In Int. Conf. Learn. Representations, 2017.

[33] S. Gu and L. Rigazio. Towards deep neural network architectures robust to adversarial examples. arXiv:1412.5068, 2014.

[34] U. Shaham, Y. Yamada, and S. Negahban. Understanding adversarial training: Increasing local stability of neural nets through robust optimization. arXiv:1511.05432, 2015.

[35] H. Zhang, M. Cisse, Y. Dauphin, and D. Lopez-Paz. Mixup: Beyond empirical risk minimization. In Int. Conf. Learn. Representations, 2018.

[36] Y. Tokozume, Y. Ushiku, and T. Harada. Between-class learning for image classification. In IEEE Conf. Comput. Vis. Pattern Recognit., pages 5486–5494, 2018.

[37] A. Bendale and T. Boult. Towards open world recognition. In IEEE Conf. Comput. Vis. Pattern Recognit., pages 1893–1902, 2015.

[38] S. Rebuffi, A. Kolesnikov, G. Sperl, and C. Lampert. iCaRL: Incremental classifier and representation learning. In IEEE Conf. Comput. Vis. Pattern Recognit., pages 2001–2010, 2017.

[39] M. Masud, J. Gao, L. Khan, J. Han, and B. Thuraisingham. Classification and novel class detection in concept-drifting data streams under time constraints. IEEE Trans. Know. Data Eng., 23(6):859–874, 2011.

[40] Z. Li, L.-F. Cheong, S. Yang, and K.-C. Toh. Simultaneous clustering and model selection: Algorithm, theory and applications. IEEE Trans. Pattern Anal. Mach. Intell., 40(8):1964–1978, 2018.

[41] T. Mensink, J. Verbeek, F. Perronnin, and G. Csurka. Distance-based image classification: Generalizing to new classes at near-zero cost. IEEE Trans. Pattern Anal. Mach. Intell., 35(11):2624–2637, 2013.

[42] S. Guerriero, B. Caputo, and T. Mensink. Deep nearest class mean classifiers. In Worskhop Int. Conf. Learn. Representations, 2018.

[43] H.-M. Yang, X.-Y. Zhang, F. Yin, and C.-L. Liu. Robust classification with convolutional prototype learning. In IEEE Conf. Comput. Vis. Pattern Recognit., pages 3474–3482, 2018.

[44] Z. Li and D. Hoiem. Learning without forgetting. IEEE Trans. Pattern Anal. Mach. Intell., 40(12):2935–2947, 2018.

[45] T. Darrell, M. Kloft, M. Pontil, G. Ratsch, and E. Rodner. Machine learning with interdependent and non-identically distributed data. Dagstuhl Reports (Dagstuhl Seminar 15152), 5(4):18–55, 2015.

[46] B. Recht, R. Roelofs, L. Schmidt, and V. Shankar. Do CIFAR-10 classifiers generalize to CIFAR-10? arXiv:1806.00451, 2018.

[47] M. Hayat, M. Bennamoun, and S. An. Deep reconstruction models for image set classification. IEEE Trans. Pattern Anal. Mach. Intell., 37(4):713–727, 2015.

[48] N. Samsudin and A. Bradley. Nearest neighbour group-based classification. Pattern Recognition, 43(10):3458–3467, 2010.

[49] P. Sarkar and G. Nagy. Style consistent classification of isogenous patterns. IEEE Trans. Pattern Anal. Mach. Intell., 27(1):88–98, 2005.

[50] S. Veeramachaneni and G. Nagy. Analytical results on style constrained Bayesian classification of pattern fields. IEEE Trans. Pattern Anal. Mach. Intell., 29(7):1280–1285, 2007.

[51] T. Dietterich, R. Lathrop, and T. Lozano-Perez. Solving the multiple instance problem with axis-parallel rectangles. Artificial Intelligence, 89:31–71, 1997.

[52] Z.-H. Zhou, Y.-Y. Sun, and Y.-F. Li. Multi-instance learning by treating instances as non-iid samples. In Int. Conf. Mach. Learn., pages 1249–1256, 2009.

[53] R. Haralick. Decision making in context. IEEE Trans. Pattern Anal. Mach. Intell., 5(4):417–428, 1983.

[54] L. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE, 77(2):257–286, 1989.

[55] J. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In Int. Conf. Mach. Learn., pages 282–289, 2001.

[56] I. Sutskever, O. Vinyals, and Q. Le. Sequence to sequence learning with neural networks. In Advances Neural Inf. Process. Syst., pages 3104–3112, 2014.

[57] T. Kipf and M. Welling. Semi-supervised classification with graph

convolutional networks. In Int. Conf. Learn. Representations, 2017.

[58] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio. Graph attention networks. In Int. Conf. Learn. Representations, 2018.

[59] I. Tsochantaridis, T. Joachims, T. Hofmann, and Y. Altun. Large margin methods for structured and interdependent output variables. J. Mach. Learn. Res., 6:1453–1484, 2005.

[60] N. Courty, R. Flamary, D. Tuia, and A. Rakotomamonjy. Optimal transport for domain adaptation. IEEE Trans. Pattern Anal. Mach. Intell., 39(9):1853–1865, 2017.

[61] X.-Y. Zhang and C.-L. Liu. Writer adaptation with style transfer mapping. IEEE Trans. Pattern Anal. Mach. Intell., 35(7):1773–1787, 2013.

[62] H. Shimodaira. Improving predictive inference under covariate shift by weighting the log-likelihood function. J. Stat. Plan. Infer., 90(2):227–244, 2000.

[63] J. Zhang, W. Li, and P. Ogunbona. Transfer learning for cross-dataset recognition: A survey. arXiv:1705.04396, 2017.

[64] Y. Ganin and V. Lempitsky. Unsupervised domain adaptation by backpropagation. In Int. Conf. Mach. Learn., pages 1180–1189, 2015.

[65] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell. Adversarial discriminative domain adaptation. In IEEE Conf. Comput. Vis. Pattern Recognit., pages 7167–7176, 2017.

[66] Y. Mansour, M. Mohri, and A. Rostamizadeh. Domain adaptation with multiple sources. In Advances Neural Inf. Process. Syst., pages 1041–1048, 2009.

[67] J. Hoffman, B. Kulis, T. Darrell, and K. Saenko. Discovering latent domains for multisource domain adaptation. In European Conf. Computer Vision, pages 702–715, 2012.

[68] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell. DeCAF: A deep convolutional activation feature for generic visual recognition. In Int. Conf. Mach. Learn., pages 647–655, 2014.

[69] H. Azizpour, A. Razavian, J. Sullivan, A. Maki, and S. Carlsson. Factors of transferability for a generic ConvNet representation. IEEE Trans. Pattern Anal. Mach. Intell., 38(9):1790–1802, 2016.

[70] R. Caruana. Multitask learning. Machine Learning, 28(1):41–75, 1997.

[71] J. Cao, Y. Li, and Z. Zhang. Partially shared multi-task convolutional neural network with local constraint for face attribute learning. In IEEE Conf. Comput. Vis. Pattern Recognit., pages 4290–4299, 2018.

[72] I. Misra, A. Shrivastava, A. Gupta, and M. Hebert. Cross-stitch networks for multi-task learning. In IEEE Conf. Comput. Vis. Pattern Recognit., pages 3994–4003, 2016.

[73] A. Zamir, A. Sax, W. Shen, L. Guibas, J. Malik, and S. Savarese. Taskonomy: Disentangling task transfer learning. In IEEE Conf. Comput. Vis. Pattern Recognit., pages 3712–3722, 2018.

[74] A. Kendall, Y. Gal, and R. Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In IEEE Conf. Comput. Vis. Pattern Recognit., pages 7482–7491, 2018.

[75] Z. Chen, V. Badrinarayanan, C.-Y. Lee, and A. Rabinovich. GradNorm: Gradient normalization for adaptive loss balancing in deep multitask networks. In Int. Conf. Mach. Learn., pages 1–10, 2018.

[76] K. Liu, Y. Li, N. Xu, and P. Natarajan. Learn to combine modalities in multimodal deep learning. arXiv:1805.11730, 2018.

[77] T. Ho, J. Hull, and S. Srihari. Decision combination in multiple classifier systems. IEEE Trans. Pattern Anal. Mach. Intell., 16(1):66–75, 1994.

[78] D. Ramachandram and G. Taylor. Deep multimodal learning: A survey on recent advances and trends. IEEE Signal Processing Magazine, 34(6):96–108, 2017.

[79] N. Neverova, C. Wolf, G. Taylor, and F. Nebout. ModDrop: adaptive multi-modal gesture recognition. IEEE Trans. Pattern Anal. Mach. Intell., 38(8):1692–1706, 2016.

[80] T. Baltrusaitis, C. Ahuja, and L. Morency. Multimodal machine learning: A survey and taxonomy. IEEE Trans. Pattern Anal. Mach. Intell., 41(2):423–443, 2019.

[81] J. Gu, J. Cai, S. Joty, L. Niu, and G. Wang. Look, imagine and match: Improving textual-visual cross-modal retrieval with generative models. In IEEE Conf. Comput. Vis. Pattern Recognit., pages 7181–7189, 2018.

[82] Y. Zhu, R. Kiros, R. Zemel, R. Salakhutdinov, R. Urtasun, A. Torralba, and S. Fidler. Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. In Int. Conf. Comput. Vis., pages 19–27, 2015.

[83] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhutdinov, R. Zemel, and Y. Bengio. Show, attend and tell: Neural image caption generation with visual attention. In Int. Conf. Mach. Learn., pages 2048–2057, 2015.

[84] L. Kaiser, A. Gomez, N. Shazeer, A. Vaswani, N. Parmar, L. Jones, and J. Uszkoreit. One model to learn them all. arXiv:1706.05137, 2017.

[85] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals. Understanding deep learning requires rethinking generalization. In Int. Conf. Learn. Representations, 2017.

[86] J. Brooks. Support vector machines with the ramp loss and the hard margin loss. Operations Research, 59(2):467–479, 2011.

[87] Y. Wu and Y. Liu. Robust truncated hinge loss support vector machines. J. American Statistical Association, 102(479):974–983, 2007.

[88] A. Ghosh, N. Manwani, and P. Sastry. Making risk minimization tolerant to label noise. Neurocomputing, 160:93–107, 2015.

[89] G. Patrini, A. Rozza, A. Menon, R. Nock, and L. Qu. Making deep neural networks robust to label noise: A loss correction approach. In IEEE Conf. Comput. Vis. Pattern Recognit., pages 1944–1952, 2017.

[90] C. Brodley and M. Friedl. Identifying mislabeled training data. J. Artif. Intell. Res., 11:131–167, 1999.

[91] L. Xu, K. Crammer, and D. Schuurmans. Robust support vector machine training via convex outlier ablation. In AAAI Conf. Artif. Intell., pages 536–542, 2006.

[92] T. Liu and D. Tao. Classification with noisy labels by importance reweighting. IEEE Trans. Pattern Anal. Mach. Intell., 38(3):447–461, 2016.

[93] D. Tanaka, D. Ikami, T. Yamasaki, and K. Aizawa. Joint optimization framework for learning with noisy labels. In IEEE Conf. Comput. Vis. Pattern Recognit., pages 5552–5560, 2018.

[94] Y. Bengio, A. Courville, and P. Vincent. Representation learning: A review and new perspectives. IEEE Trans. Pattern Anal. Mach. Intell., 35(8):1798–1828, 2013.

[95] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol. Extracting and composing robust features with denoising autoencoders. In Int. Conf. Mach. Learn., pages 1096–1103, 2008.

[96] J. Yang, D. Parikh, and D. Batra. Joint unsupervised learning of deep representations and image clusters. In IEEE Conf. Comput. Vis. Pattern Recognit., pages 5147–5156, 2016.

[97] Z. Wu, Y. Xiong, S. Yu, and D. Lin. Unsupervised feature learning via non-parametric instance discrimination. In IEEE Conf. Comput. Vis. Pattern Recognit., pages 3733–3742, 2018.

[98] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick. Momentum contrast for unsupervised visual representation learning. arXiv:1911.05722, 2019.

[99] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. BERT: Pretraining of deep bidirectional transformers for language understanding. arXiv:1810.04805, 2018.

[100] C. Doersch and A. Zisserman. Multi-task self-supervised visual learning. In Int. Conf. Comput. Vis., pages 2051–2060, 2017.

[101] A. Rasmus, H. Valpola, M. Honkala, M. Berglund, and T. Raiko. Semi-supervised learning with ladder networks. In Advances Neural Inf. Process. Syst., pages 3546–3554, 2015.

[102] D.-H. Lee. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In Workshop Int. Conf. Mach. Learn., 2013.

[103] A. Blum and T. Mitchell. Combining labeled and unlabeled data with co-

training. In Annual conf. Comput. Learn. Theory, pages 92–100, 1998.
[104] Z.-H. Zhou and M. Li. Tri-training: exploiting unlabeled data using three classifiers. IEEE Trans. Know. Data Eng., 17(11):1529–1541, 2005.
[105] J. Springenberg. Unsupervised and semi-supervised learning with categorical generative adversarial networks. In Int. Conf. Learn. Representations, 2016.
[106] Z. Dai, Z. Yang, F. Yang, W. Cohen, and R. Salakhutdinov. Good semi-supervised learning that requires a bad GAN. In Advances Neural Inf. Process. Syst., pages 6510–6520, 2017.
[107] M. Sajjadi, M. Javanmardi, and T. Tasdizen. Regularization with stochastic transformations and perturbations for deep semi-supervised learning. In Advances Neural Inf. Process. Syst., pages 1163–1171, 2016.
[108] S. Laine and T. Aila. Temporal ensembling for semi-supervised learning. In Int. Conf. Learn. Representations, 2017.
[109] A. Tarvainen and H. Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In Advances Neural Inf. Process. Syst., pages 1195–1204, 2017.
[110] B. Lake, R. Salakhutdinov, and J. Tenenbaum. Human-level concept learning through probabilistic program induction. Science, 350(6266):1332–1338, 2015.
[111] C. Lampert, H. Nickisch, and S. Harmeling. Attribute-based classification for zero-shot visual object categorization. IEEE Trans. Pattern Anal. Mach. Intell., 36(3):453–465, 2014.

【作者简介】



张煦尧，中国科学院自动化研究所模式识别国家重点实验室副研究员，中国自动化学会模式识别与机器智能专委会副秘书长。主要研究兴趣包括：模式识别、机器学习、文字识别、深度学习等。



刘成林，中国科学院自动化研究所副所长，模式识别国家重点实验室主任，中国人工智能学会副理事长，中国自动化学会模式识别与机器智能专委会主任，IEEE/IAPR/CAAI Fellow。主要研究兴趣包括：模式识别、图像处理、神经网络、机器学习、文档图像分析与识别等。